NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Technical Report 32-1366

Optimal Nonlinear Estimation Based on Orthogonal Expansions

William Kizner

JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA

April 15, 1969

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Technical Report 32-1366

Optimal Nonlinear Estimation Based on Orthogonal Expansions

William Kizner

JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA

April 15, 1969

TECHNICAL REPORT 32-1366

Prepared Under Contract No. NAS 7-100
National Aeronautics and Space Administration

Preface

The work described in this report was performed by the Systems Division of the Jet Propulsion Laboratory.

Contents

Tables

Figure

Abstract

This report is concerned with the statistical optimal estimation of parameters x of nonlinear systems. If a quadratic loss function is used, then the conditional mean of x given the observations and *a priori* information is used. The conditional mean is calculated using an n-dimensional Gram–Charlier expansion of the *a posteriori* probability density function. A numerical method for finding this distribution is given, and some of its properties are discussed, including the case for which the method is optimal.

If a more general kind of loss function is used and a mean-square approximation to the *a posteriori* probability density function is desired, a Hermite function expansion which converges in the mean square is employed. Again, numerical methods are given for finding the coefficients of the expansion, and the optimality of this procedure is discussed.

Optimal Nonlinear Estimation Based on Orthogonal Expansions

I. Introduction

This paper is concerned with the statistical optimal estimation of parameters of nonlinear systems. The methods presented herein are being applied to various problems in orbit determination to ascertain the effects of nonlinearities in the system.

Let x be an $n \times 1$ vector of parameters to be estimated. Although the methods given here do not require much more computation than for maximum likelihood methods when n is 1 (in fact, the amount of computation required may be less), the amount of computation increases rapidly with n. Thus for large n these methods will probably not be used routinely, but as checks on other, simpler methods, such as maximum likelihood estimates. These methods can serve not only as a check on the suitability of using a maximum likelihood estimate but also to ascertain whether a purported maximum likelihood estimate is just that. Here, certain partial derivatives involved in maximum likelihood estimates which may be subject to error are not used.

Let z be the $k \times 1$ vector of observations and y be the $k \times 1$ vector of observations which would result from a noiseless system. For simplicity, assume that x is a constant of the time. Also,

$$y = v(x) \tag{1}$$

Let $p(x \mid z)$ denote the *a posteriori* probability density function of x given z, $p(x)$ the *a priori* density function of x, and $P(z \mid x)$ the probability density for z given x. Then Bayes' rule (Ref. 1) implies that

$$p(x \mid z) = \frac{P(z \mid x)\, p(x)}{\int_{-\infty}^{\infty} P(z \mid x)\, p(x)\, dx} \tag{2}$$

It will be assumed that $P(z \mid x)$ and $p(x)$ can be calculated to within an unknown constant. Let

$$L(z, x) = k\, P(z \mid x)\, p(x) \tag{3}$$

be the likelihood function including *a priori* information and k an unknown constant > 0. The function $L(z, x)$ is the likelihood function which is maximized with respect to z when a maximum likelihood estimate that includes *a priori* information is being obtained. Then

$$p(x \mid z) = \frac{L(z, x)}{\int L(z, x)\, dx} \tag{4}$$

In the event that the loss function is quadratic, it is well known that the conditional parameter estimate

$$x^* = \int_{-\infty}^{\infty} x\, p(x \mid z)\, dx \tag{5}$$

is statistically optimum. It will be shown that the method based on the n-dimensional Gram–Charlier expansion (Ref. 2) is the indicated approach. A representation that is close to this approach is based on the n-dimensional Edgeworth series or quasimoment functions (Ref. 3), a rearrangement of the Gram–Charlier series that normally improves the convergence with respect to the uniform and mean-square norms. However, the n-dimensional Gram–Charlier expansion may not represent any probability density function in the mean-square sense even if the function is very well-behaved and is square integrable. Examples of this will be given, and the explanation for the failure of the convergence will be discussed. If the cost function is not quadratic and an approximation of $p(x \mid z)$ is desired which is close in the mean-square sense, a different approach based on the best mean-square approximation is used (Ref. 4).

II. Properties of Two Expansions

The dimension of the probability density distributions will be taken equal to 1. The generalization to n dimensions is straightforward. We start with the Gram–Charlier expansion.

Let the density function $p(x)$ be represented by

$$p(x) \approx \frac{1}{2\pi} e^{-x^2/2} \left[\sum_{i=0}^{\infty} a_j \, He_j \, (x) \right] \tag{6a}$$

$$a_j = \frac{1}{j!} \int_{-\infty}^{\infty} p(x) \, He_j(x) dx \qquad j = 0, 1, \cdots \tag{6b}$$

where $He_j(x)$ is the Hermite polynomial of jth degree, and $a_0 = 1$. Explicitly,

$$He_0(x) \equiv 1$$

$$He_1(x) \equiv x$$

and

$$He_{j+1}(x) \equiv xHe_j(x) - jHe_{j-1}(x) \qquad j = 1, 2, \cdots \tag{7}$$

It is known (Ref 2) that the $He_j(x)$ are mutually orthogonal with respect to the weight function $e^{-x^2/2}$.

From the definition of $He_j(x)$ (7) and (6), a_1 is the mean of the probability density function and

$$2a_2 = E(x^2) - 1 \tag{8}$$

where E is the expectation operator. Thus the mean and variance of the density function (if they exist) are determined from a_1 and a_2 regardless of how poor an approximation the series (6) is to $p(x)$. Likewise, it can be shown that moments up to mth order (if they exist) can be determined exactly from a knowledge of the coefficients up to a_m, which are sometimes called the quasimoments.

Conditions for the pointwise convergence of this series have been given by Cramer (Ref. 2, p. 223). Most of the important distributions used in statistics do not satisfy these conditions.

It is worthwhile to examine the reasoning which has led some to conclude that the Gram–Charlier series, or the quasimoment function series, converges in the mean-square sense. It is well known that the functions of the form

$$\sum_{j=0}^{m} e^{-x^2/2} \, a_j \, He_j \, (x) \tag{9}$$

are everywhere dense in $L_2 \, (-\infty, \infty)$. Since all probability density functions used in practice belong to $L_2 \, (-\infty, \infty)$ (they are bounded), one is tempted to conclude that the expansion converges in the mean square.

The fallacy in the argument is that there indeed is a sequence of functions of the form (9) which converges in the mean, but that in general the coefficients a_j are functions of m.

Thus an expansion which is best in the mean-square sense for each partial sum is desirable. Such an expansion is given by (Ref. 5)

$$p(x) \approx \sum_{j=0}^{\infty} b_j (2^j j! \pi^{1/2})^{-1/2} \, e^{-x^2/2} \, H_j(x) \tag{10a}$$

$$b_j = \int_{-\infty}^{\infty} p(x) \, (2^j j! \pi^{1/2})^{-1/2} \, e^{-x^2/2} \, H_j(x) \tag{10b}$$

where the $H_j(x)$ are also called Hermite polynomials and are given by

$$H_0(x) \equiv 1$$

$$H_1(x) \equiv 2x$$

$$H_{j+1}(x) \equiv 2xH_j(x) - 2jH_{j-1}(x) \tag{11}$$

This expansion is called here the Hermite series as opposed to the Gram–Charlier series.

It can be shown that the terms $e^{-x^2/2}H_j(x)$ are mutually orthogonal in $L_2\,(-\infty,\,\infty)$ (with unit weight function). As a consequence (10) is the best approximation in the mean-square sense for $p(x)$.

The following theorem is proved by Sansone (Ref. 5, Section 4.10).

Theorem 1. Let $f(x)$ be continuous, of bounded variation, and $\epsilon L_1(-\infty,\,\infty)\cap L_2(-\infty,\,\infty)$. Then the series (10), with $p(x)$ replaced by $f(x)$, is uniformly convergent in any interval interior to $(-\infty,\,\infty)$.

Schwartz (Ref. 4) has proved the following two theorems on the rate of decrease of the coefficients b_j in (10).

Theorem 2. Assume that the derivative of $f(x)$ exists and that the function $[xf(x) - f'(x)]\ \epsilon L_2(-\infty,\,\infty)$. Then the coefficients b_j, $j = 1, 2, \cdots$, satisfy the bound

$$|\,b_j\,| < \frac{c}{(2j)^{1/2}} \tag{12}$$

where c is the $L_2(-\infty,\,\infty)$ norm of $[xf(x) - f'(x)]$.

Theorem 3. Assume that the function

$$e^{x^2/2}\,\frac{d^r}{dx^r}\left[\,e^{x^2/2}\,f(x)\right] = \sum_{i=0}^{r}\left[\frac{r!}{i!(r-i)!}\right](-1)^i\,2^{-i/2}H_i(x/2^{1/2})\,\frac{d^{r-i}}{dx^{r-1}}\,f(x) \tag{13}$$

exists and is square integrable. Then the coefficients b_j, $j = 1, 2, \cdots$, are bounded by

$$|\,b_j\,| < \frac{c(r)}{(2j)^{r/2}} \tag{14}$$

where $c(r)$ is the $L_2(-\infty,\,\infty)$ norm of (13).

Comment. The L_2 assumption on the function in (13) can be replaced by an L_1 requirement. Thus the existence of high-order moments and the existence and integrability of high-order derivatives of a probability density function insure the rapid convergence of the Hermite series.

Another question which may arise is whether the series (10) is convergent on the complete interval $(-\infty,\,\infty)$. Let $\phi_j(x)$ denote the normalized Hermite function

$$\phi_j(x) = (2^j j!\pi^{1/2})^{-1/2}\,e^{-x^2/2}\,H_j(x) \tag{15}$$

Thus functions are known to be uniformly bounded on $(-\infty,\,\infty)$. Professor Schwartz has pointed out (in a private communication) that if $\sum_{j=0}^{\infty} b_j$ is absolutely convergent, then since $\lim_{x\to\pm\infty}\phi_j(x) = 0$, the series is equal to zero at $\pm\infty$. Thus convergence is assured if $\lim_{x\to\pm\infty} f(x) = 0$. Moreover, it is easy to see that convergence is then uniform on $(-\infty,\,\infty)$. The series will be absolutely convergent if $f(x)$ satisfies theorem 3 with $r \geq 3$.

III. The Numerical Evaluation of the Coefficients of the Hermite Series

Let $g(x) = e^{-x^2/2}\,\pi_m(x)$, where $\pi_m(x)$ is a polynomial in x of degree m. Let x_ν^N, $\nu = 1, 2, \cdots, N$, denote the N zeroes of $H_N(x)$, and W_ν^N the weights for Gaussian quadrature that are chosen so that

$$\int_{-\infty}^{\infty}\pi_m(x)\,e^{-x^2}\,dx = \sum_{\nu=1}^{N} W_\nu^N\pi_m(x_\nu^N) \tag{16}$$

whenever $2N-1 \geq m$.

The most extensive tabulation of the zeroes and weights is by Stroud and Secrest (Ref. 6). Then we have:

Theorem 4. Let $g(x) = e^{-x^2/2}\pi_m(x)$, where $\pi_m(x)$ is a polynomial in x of degree m or less. Then

$$\int_{-\infty}^{\infty} g(x)\,e^{-x^2/2}\,H_j(x)\,dx = \sum_{\nu=1}^{N} g(x_\nu^N)\,\exp\,(x_\nu^2\,/2)\,H_j\,(x_\nu^N)\,W_\nu^N \tag{17}$$

provided N is chosen so that $m + j \leq 2N-1$.

Proof. We write the integral in the form

$$\int_{-\infty}^{\infty}[g(x)\,e^{x^2/2}\,H_j(x)]\,e^{-x^2}\,dx \tag{18}$$

and note that the quantity in the brackets is a polynomial of degree $j + m$ at most. Then (17) follows from the result on Gaussian quadrature.

This preceding formula suggests that we choose as an approximation for b_j the following:

$$b_j^N = (2^j j! \pi^{1/2})^{-1/2} \sum_{\nu=0}^{N} p(x_\nu^N)\, H_j(x_\nu^N) \exp(x_\nu^2/2)\, W_\nu^N \quad \text{(19a)}$$

and

$$p^N(x) = \sum_{\nu=0}^{N-1} b_j^N\, \phi_j(x) \quad \text{(19b)}$$

Let $x_1, x_2, \cdots, x_N$ be N distinct values of x, and:

Lemma. Let ϕ_j (see Eq. 15) be a vector with n components $\phi_j(x_1)$, $\phi_j(x_2)$, $\cdots$, $\phi_j(x_N)$. Then the vectors $\phi_0, \phi_1, \cdots, \phi_{N-1}$ are linearly independent.

Proof. It is well known that the polynomials H_j considered as vectors with elements $H_j(x_1)$, $H_j(x_2)$, $\cdots$, $H_j(x_N)$ are linearly independent. Then the array of vectors ϕ_j, $j = 0, 1, \cdots, N-1$ are independent since it can be formed by the matrix multiplication of a diagonal matrix with nonzero elements and the H_j.

Theorem 5. $p(x) = p^N(x)$ at x_ν^N, $\nu = 1, 2, \cdots, N$.

Proof. We can determine a unique combination of Hermite polynomials $\sum_{\nu=1}^{N-1} b_\nu H_\nu(x)$ such that $e^{-x^2/2} \sum_{\nu=1}^{N-1} b_\nu H_\nu(x) = p(x)$ at the N points x_ν^N. This follows from the previous lemma. It is clear that (10a) is an exact equality if $p(x)$ is a finite Hermite series. From (17) and (10) we can determine these coefficients exactly by (19).

Theorem 6. The approximation given by (19) is the best approximation in the mean square whenever the function to be approximated $p(x)$ is given by

$$p(x) = \sum_{i=0}^{N} b_i\, \phi_i(x) \quad \text{(20)}$$

and N interpolation points or more are used.

Proof. The approximation $p^N(x) = \sum_{i=0}^{N-1} b_i\ \phi_i(x)$ and $p(x)$ can differ only by $b_N\, \phi_N(x)$.

This result leads one to suspect that (19) is a good approximation in the mean-square sense whenever $p(x)$ has a rapidly converging Hermite series.

We now discuss an extremal property of (19) which is important when (19b) is truncated. In other words, suppose that we consider the approximation

$$p_M^N(x) = \sum_{j=0}^{M} b_j^N\ \phi_j(x) \quad \text{(19c)}$$

where $M \leq N-1$. Then we have:

Theorem 7. The approximation (19c) is the weighed least-squares approximation to $p(x)$ at the points x_ν^N, $\nu = 1, 2, \cdots, N$ with weights equal to $W_\nu^N \exp\ [(x_\nu^N)^2]$.

More exactly

$$J = \sum_{\nu=1}^{N} [p(x_\nu^N) - \Sigma c_j\, \phi_j(x_\nu^N)]^2\, W_\nu^N \exp[(x_\nu^N)^2] \quad \text{(21)}$$

is minimized with respect to the coefficients c_j if each c_j is made equal to b_j^N given in (19a).

Proof. Define the scalar product $<d, e> = \sum_{i=1}^{N} d_i e_i W_i$ where d, e and W are N-dimensional vectors. Then (19a) can be written as

$$b_j^N = \langle p(x_\nu^N), \phi_j(x_\nu^N) \rangle \quad \text{(22)}$$

where

$$W_i = W_i^N \exp[(x_i^N)^2]$$

From theorem 5 we have

$$\langle \phi_j(x_\nu^N), \phi_k(x_\nu^N) \rangle = \delta_{jk} \quad \text{(23)}$$

The conclusion of the theorem follows from a standard result of Hilbert space theory when p_M^N is viewed as a projection of $p(x_\nu)$ on the $M+1$ dimensional subspace generated by the ϕs.

It is of interest to tabulate the weights for various values of $N(4, 8, 16)$ (Table 1). In column 1 of this table we give the value of x_ν, starting with the smallest in absolute value. The negative values of x_ν are not given because of symmetry. In the third column, the weights for the scalar product $(W_\nu^N \exp(x_\nu^{N2}))$ are given. We next show that these weights are approximately proportional to the distance between the x_ν^N.

Table 1. Relationship between scalar product weights and interval lengths

Interpolation point	Intermediate point	$W_i^N \exp(x_i^N)^2 = W_i$	W_i/interval length
$N = 4$ 0.52465 1.65068	0 1.04930 2.25207	1.05996 1.24023	1.01017 1.03114
$N = 8$ 0.38119 1.15719 1.98166 2.93064	0 0.76237 1.55201 2.41130 3.44997	0.76454 0.79289 0.86675 1.07193	1.00285 1.00412 1.00869 1.03202
$N = 16$ 0.27348 0.82295 1.38026 1.95179 2.54620 3.17700 3.86945 4.68874	0 0.54696 1.09894 1.66158 2.24200 3.50359 3.50359 4.23530 5.14218	0.54738 0.55244 0.56322 0.58125 0.60974 0.65576 0.73825 0.93687	1.00076 1.00084 1.00104 1.00142 1.00219 1.00393 1.00893 1.03308

We now make precise the manner in which we will compare the discrete approximation for the Fourier coefficients (19a) to the continuous one (10b). We use the quadrature formula

$$\int_{x_0}^{x_0+h} f(x)dx = h\,f\left(x_0 + \frac{h}{2}\right) \tag{24}$$

which has a truncation error of about half the error of the trapezoidal rule. We now find "intermediate points" y_ν^N $\nu = 1, 2, \cdots, N+1$ such that each x_ν^N is in the center of the interval $[y_\nu^N, y_{\nu+1}^N]$. These points are given in column 2 of Table 1. Thus we can associate an interval length $y_{\nu+1}^N - y_\nu^N$ with each x_ν^N. The ratio of the weights (in column 3) divided by the interval length is given in column 4. It can be seen that this ratio is approximately equal to 1. Thus if $\int p^2(x)dx$ were zero outside the intervals that we cover here, (19a) would be a rough numerical approximation to (10b).

However, more can be said about this approximation as $N \to \infty$. We consider the values in column 4 which we associate with $x = 1$, 2, 3, and 4. Thus for $x = 1$ and $N = 16$ we associate the first interval shown with this x, and the ratio is 1.00084. In Fig. 1, we show how this ratio minus 1 approaches zero with increasing N.

If $p(x) \in L_2(-\infty, \infty)$ then $\int p^2(x)dx$ is essentially zero outside some finite interval (a, b). Within this interval

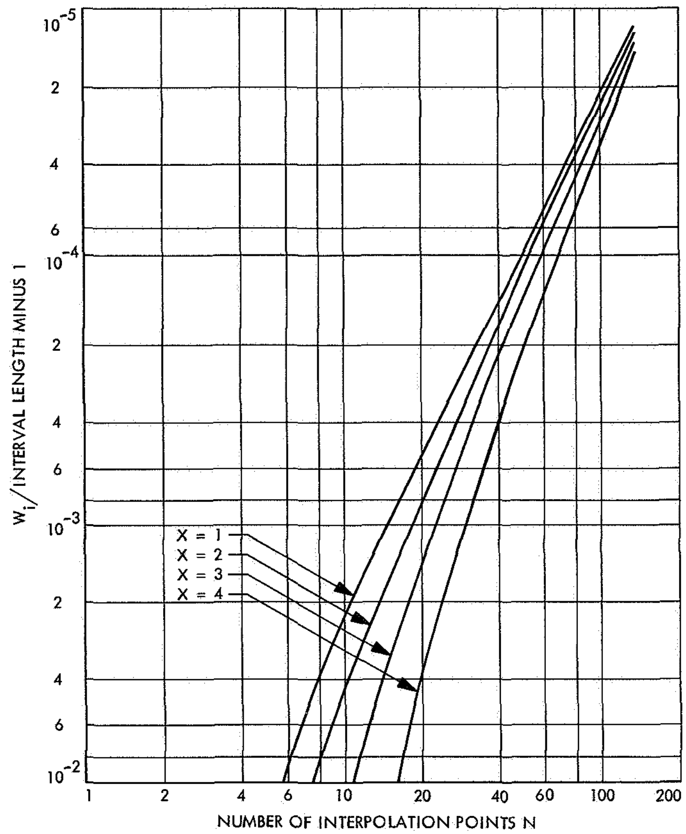

Fig. 1. Numerical data to support conjecture that $\lim_{N\to\infty} a_j^N = a_j$

our numerical approximation for the Fourier coefficients seems to approach the exact value, provided the integral in (10b) exists as a Riemann integral. Thus it is not unreasonable to state the following:

Conjecture.

$$\lim_{N\to\infty} a_j^N = a_j \tag{25}$$

whenever the integral in (10b) exists as a Riemann integral.

Later we will need to calculate the area and moments of the unnormalized probability distribution. By making use of the recurrence relation (11) and the relation for the derivative

$$H_j'(x) = 2jH_{j-1}(x) \tag{26}$$

$$\int_a^b e^{-x^2/2} H_j(x)dx = -2\left[H_j(x)e^{-x^2/2}\right]_a^b + 2j\int_a^b H_{j-1}(x)e^{-x^2/2}\,dx \tag{27}$$

It easily follows that

$$\int_{-\infty}^{\infty} e^{-x^2/2} H_j(x)\, dx = (2\pi)^{1/2} \frac{j!}{\left(\frac{j}{2}\right)!} \qquad j = 0,2,4, \cdots$$

$$= 0 \qquad j = 1,3,5 \cdots \tag{28}$$

Thus one method which suggests itself for approximating an unnormalized probability density function is to use (28) to evaluate the area of the approximation $p_M^N(x)$, and to divide the coefficients b_i by this factor, which will make the resulting area equal to 1. The question now arises as to whether this sequence will converge in the mean square to the probability density function, since the best mean-square approximations for this function may not have unit area. A partial answer is provided by theorem 9, which depends in part on an extrapolation of numerical data. This theorem concerns the nature of the oscillations of $\phi_n(x)$, which are roughly sinusoidal, and the area $\int_R^\infty \phi_n(x)dx$, where R is large. It has been found empirically that for $R > 0$, $\left|\int_R^\infty \phi_n(x)dx\right|$ considered as a function of R assumes its greatest value for R equal to the largest zero of $\phi_n(x)$. Call the resulting value of the integral R_n. Also, the ratio of R_n to $\int_0^\infty \phi_n(x)dx$ appears to approach a limit as n tends to infinity or at least to decrease very slowly. For $n = 88, 96, 104, 120$, and 136, $R_n/\int_0^\infty \phi_n(x)dx$ is 1.27982, 1.27950, 1.27922, 1.27878, and 1.27842. At any rate it seems reasonable to assume that for $n \geq 88$ the ratio is less than 1.28.

Another result that will be needed is the simplification in evaluating $\int_{-\infty}^{\infty} b_n\, \phi_n(x)dx$ based on (28). It can easily be shown that

$$\int_{-\infty}^{\infty} \phi_n(x)dx = \left(\frac{n-1}{n}\right)^{1/2} \int_{-\infty}^{\infty} \phi_{n-2}(x)dx$$

$$n = 2,4,6, \cdots \tag{29}$$

and

$$\int_{-\infty}^{\infty} \phi_0(x)dx = (2)^{1/2}(\pi)^{1/4} \tag{30}$$

Theorem 8. Let $p(x)$ be a true probability density function having unit area. Let the partial sums of (10) be denoted by

$$S_n(x) = \sum_{i=0}^{n} b_i\, \phi_i(x) \tag{31}$$

and the area of each of these by

$$A_n = \int_{-\infty}^{\infty} S_n(x)dx \tag{32}$$

Then $\lim_{n\to\infty} A_n = 1$ provided that $\sum_{i=0}^{\infty} b_i$ is absolutely convergent.

Proof. For a given $\epsilon > 0$ an $N(\epsilon)$ will be shown to exist that depends on the sequence such that $|A_n - 1| < \epsilon$ for $n \geq N(\epsilon)$. From the fact that $\sum_{i=0}^{\infty} b_i$ is absolutely convergent, there exists an integer K such that

$$\sum_{i=K+1}^{\infty} |b_i| \leq \frac{\epsilon}{8} \frac{1}{(2\pi^{1/4})^{1/2}} \frac{1}{1.28} \tag{33}$$

Next an $R > 0$ can be found (since p and $S_K \in L_1(-\infty, \infty)$) such that

$$\left|\int_{-\infty}^{-R} p(x)dx\right| + \left|\int_R^{\infty} p(x)dx\right| < \frac{\epsilon}{4}$$

and

$$\left|\int_{-\infty}^{-R} S_n(x)dx\right| + \left|\int_R^{\infty} S_n(x)dx\right| \leq \frac{\epsilon}{4} \tag{34}$$

Within $(-R, R)$ it is known that L_2 convergence implies L_1 convergence (here the measure of the domain of the function is finite), and so when the error of the approximation is less than $\epsilon/4$ within the interval, the total error for the area will be less than $\epsilon/4$ plus $\epsilon/4$ contributed by the tails of $\sum_{i=K+1}^{\infty} b_i\, \phi_i(x)$, plus $\epsilon/4$ from the tails of $p(x)$ and $\epsilon/4$ from the tails of $\sum_{i=0}^{K} b_i\phi_i(x)$.

Thus we have shown that as far as the Cauchy principle value of the integral of the approximation is concerned, the error in the area tends to zero. It is easy to show that under the given hypothesis the area of the approximation exists no matter which way the limit is taken. It is then a simple matter to prove.

Theorem 9. Consider the modified sequence of the partial sums of (10), $S_n(x)/A_n$ for a probability density function $p(x)$. If $\lim_{n\to\infty} A_n = 1$ then L_2 convergence to $p(x)$ occurs.

To find the moments one may use the recurrence relation (11) together with (28).

Thus we see one way of finding the area of the unnormalized probability distribution and its moments is to take the approximate solution, $p^N(x)$ defined by (19), and to use (29) and the recurrence relation to obtain the area and moments. Our experience with this method has been good, and for functions which closely resemble the Gaussian distribution the results have been better than for the following method, which can be better justified theoretically.

An alternative method of finding the area and moments is to multiply the value of $p(x)$ at the interpolation points by x, x^2, etc., and to find the area of the resulting functions. This procedure seems to work better when $p(x)$ is not approximately Gaussian.

IV. The Expansion in Hermite Functions: Finite Dimensional Case

For the case where x is of dimension k, $f(x) = f(x_1, x_2, \cdots, x_k)$ can be represented by

$$f(x) \approx \Sigma\, a_{j_1 j_2, \cdots, j_k} \phi_{j_1}(x_1)\, \phi_{j_2}(x_2), \cdots, \phi_{j_k}(x_k) \qquad (35)$$

where

$$\phi_{j_i}(x_i) = \exp(-x_i^2/2)\,(2^j j!\pi^{1/2})^{-1/2} H_{j_i}(x_i) \qquad (36)$$

provided $f(x) \epsilon L_2$ in the k-dimensional space. This follows from the well-known result that the products $\phi_{j_1}(x_1)\,\phi_{j_2}(x_2), \cdots, \phi_{j_k}(x_k)$ are orthonormal and form a complete system in the space of square integrable functions of the k-dimensional product space whenever $\phi_j(x)$ are orthogonal and complete in the one-dimensional space.

In order to evaluate the coefficients we make use of a generalization of the theorem on Gaussian quadrature for a product space, see Stroud and Secrest (Ref. 6).

Theorem 10. Let ξ_ν and W_ν be the zeroes and weights for the Gaussian quadrature in one dimension for interval (a,b). Then for polynomials $p_{j_i}(x_i)$, $i = 1,2, \cdots, k$

$$\int_a^b \int_a^b, \cdots, \int_a^b p_{j_1}(x_1) p_{j_2}(x_2) \cdots p_{j_k}(x_k) w(x_1) w(x_2) \cdots w(x_k) dx_1 dx_2 \cdots dx_k$$

$$= \sum_{\nu_1=1}^{N} \sum_{\nu_2=1}^{N} \cdots \sum_{\nu_k=1}^{N} p_{j_1}(\xi_{\nu_1}) p_{j_2}(\xi_{\nu_2}) \cdots p_k \times (\xi_{\nu_k}) W_{\nu_1} W_{\nu_2} \cdots W_{\nu_k} \qquad (37)$$

whenever the degree j_i of each polynomial $p_{j_i}(x_i)$ is less than or equal to $2N$-1.

The theorem can be proved by mathematical induction by writing the integral in iterated form. If now we define

$$\overline{W}_j = W_j \exp(x_j^2) \qquad (38)$$

Then

$$a^N_{j_1 j_2 \cdots j_k} = \sum_{\nu_1=1}^{N} \sum_{\nu_2=1}^{N} \cdots \sum_{\nu_k=1}^{N} f(\xi_{\nu_1}, \xi_{\nu_2} \cdots \xi_{\nu_k}) \times \phi_{j_1}(\xi_{\nu_1}) \phi_{j_2}(\xi_{\nu_2}) \cdots \phi_{j_k}(\xi_{\nu_k}) \overline{W}_{\nu_1} \overline{W}_{\nu_2} \cdots \overline{W}_{\nu_k} \qquad (39)$$

V. An Example of the Use of the Method Based on the Hermite Expansion

Suppose we wish to estimate θ in a nonlinear function $f(\theta,t) = 100 \sin(t + \theta)$ and have two observations y_i made at $t = 0°$ and $t = 1°$ of 87.6025 and 86.4307. We assume that the errors of the two observations are independent Gaussian variables with zero mean and unit variance. Then it can be shown that at least one local minimum for the sum of the squares of the residuals is at $\theta = 60°$.

Let

$$A = \left.\frac{\partial f}{\partial \theta}\right|_{t=t_i} = \begin{bmatrix} 100 \cos(\theta) \\ 100 \cos(\theta + 1°) \end{bmatrix}$$

where the derivatives are with respect to a change in θ in radians. Then it can be verified that the normal equations

$$(A^T A)\, \Delta\theta = A_T(y - f)$$

are satisfied for $\Delta\theta = 0$. $(A^TA)^{-1}$ is an estimate of the variance of the MLE estimate in radians. The resulting standard deviation is 0.822685° for the estimate for θ.

We now expand the conditional probability density by (10), using the variable

$$x = \frac{\theta - 60°}{0.822685°}$$

To do this, we evaluate $p(y|x)$ at the selected points given by the zeroes of $H_n(x)$. If we do not know the *a priori* probability density we can set that factor equal to 1. The $p(x)\ p(y|x)$ is proportional to

$$\exp\left\{-\frac{1}{2}\left\{\sum_{i=1}^{2}\left[y_i - f(\theta, t_i)\right]^2\right.\right.$$

where $\theta = x(0.822685°) + 60$.

We list the approximate area and moments in Table 2 for various N. It turns out that the conditional probability density is noticeably skewed and that the expected value of x is about 0.03834. Thus the nonlinear estimate differs from the maximum-likelihood estimate by an amount of about 4% of the estimated standard deviation.

Table 2. Convergence of Hermite series

Name of distribution	Equation	Remarks	First 10 Fourier coefficients	Scale factor	Degree of polynomial for reference	N or number of interpolation points	Area of unnormalized distribution	Mean	Variance	$L_2(-\infty,\infty)$ norm of distribution	$L_\infty(-\infty,\infty)$ norm of distribution
Unknown phase angle problem	Described in Section V		0.47462	1	59	72	0.89464	0.03834	1.00607	0.47470	0.35690
			0.00003							L_2 norm	L_∞ norm
			0.00640							of error	of error
			0.00013			2	0.89351	0.00636	0.99996	0.00759	0.00512
			0.00038			3	0.89378	0.01904	1.00060	0.00589	0.00362
			0.00002			4	0.89402	0.03808	1.00034	0.00055	0.00028
			0.00524			5	0.89436	0.03811	1.00296	0.00026	0.00015
			0.00000			6	0.89450	0.03818	1.00417	0.00017	0.00012
			0.00033			7	0.89464	0.03823	1.00599	0.00002	0.00002
			0.00000			8	0.89464	0.03830	1.00600	0.00001	0.00001
						9	0.89464	0.03832	1.00603	0.00001	0.00000
						10	0.89464	0.03834	1.00604	0.00000	0.00000
						12	0.89464	0.03834	1.00607	0.00000	0.00000
Cauchy distribution	$\frac{1}{\pi(1+x^2)}$	No moments exist; odd Fourier coefficients are zero and are not given	0.39296	1	40	72	1[a]	–	–	0.39883	0.31830
			0.00974								
			0.01397								
			0.01316			2	0.68300			0.15409	0.09920
			0.05602			3	0.71639			0.18979	0.12916
			0.00757			4	0.75308			0.13865	0.08899
			0.01267			5	0.81214			0.10168	0.07385
			0.00880			6	0.80872			0.09811	0.06326
			0.02313			7	0.81753			0.09811	0.07105
			0.00605			8	0.83127			0.08649	0.05736
						9	0.84933			0.07543	0.05647
						10	0.85155			0.07339	0.04999
						12	0.86382			0.06661	0.04569
						14	0.87482			0.06001	0.04309
						16	0.88278			0.05561	0.03951
						20	0.89558			0.04842	0.03557
						24	0.90497			0.04327	0.03252
						28	0.91223			0.03936	0.03006
						32	0.91806			0.03625	0.02804
						36	0.92287			0.03371	0.02635
						40	0.92694			0.03157	0.02491

Table 2 (contd)

Name of distribution	Equation	Remarks	First 10 Fourier coefficients	Scale factor	Degree of polynomial for reference	*N* or number of interpolation points	Area of unnormalized distribution	Mean	Variance	$L_2(-\infty,\infty)$ norm of distribution	$L_\infty(-\infty,\infty)$ norm of distribution
Normalized	$\frac{1}{(\nu\pi)^{1/2}}\frac{\Gamma[(\nu+1)/2]}{\Gamma(\nu/2)}$	3rd and	0.59931	$(3)^{1/2}$	42	72	1[a]	0[a]	1[a]	0.63078	0.63662
student *t*	$\cdot\left(1+\frac{x^2}{2}\right)^{-(\nu+1)/2}$	higher	−0.01955								
distribution,		moments	−0.15988								
$\nu = 3$		do not	0.01488			2	0.91067	0	1.00000	0.20823	0.23768
		exist. Odd	0.09454			3	0.89227	0	0.57690	0.22725	0.15721
		Fourier	−0.00923			4	0.92577	0	0.40307	0.12473	0.11861
		coeffi-	−0.04784			5	1.07668	0	1.29662	0.08371	0.05637
		cients	0.00747			6	0.98920	0	0.80793	0.07643	0.08905
		are zero	0.03347			7	0.95940	0	0.70350	0.07540	0.05155
		and are	−0.00476			8	0.98238	0	0.59420	0.05122	0.05219
		not given.				9	1.02115	0	1.13392	0.03914	0.02778
		The prob-				10	0.99623	0	0.79029	0.03452	0.03988
		lem is				12	0.99305	0	0.69060	0.02477	0.02545
		scaled so				14	0.99771	0	0.79614	0.01755	0.01979
		that the				16	0.99634	0	0.74554	0.01317	0.01332
		variance				20	0.99770	0	0.78023	0.00748	0.00783
		is 1.				24	0.99839	0	0.80398	0.00446	0.00492
						28	0.99879	0	0.82130	0.00276	0.00321
						32	0.99904	0	0.83457	0.00176	0.00217
						36	0.99922	0	0.84513	0.00145	0.00151
						40	0.99934	0	0.85380	0.00075	0.00109
Normalized	Same	20th and	0.53658	1.0540	48	72	1[a]	0[a]	1[a]	0.53685	0.41530
student *t*		higher	−0.00002								
distribution,		moments	−0.01497								
$\nu = 20$		do not	0.00005			2	1.00250	0	1.00000	0.01803	0.01636
		exist. Odd	0.00841			3	0.98099	0	0.87764	0.01588	0.00896
		Fourier	0.00000			4	0.99106	0	0.92322	0.00970	0.00657
		coeffi-	−0.00053			5	1.00048	0	1.00259	0.00104	0.00066
		cients are	0.00001			6	0.99963	0	0.99455	0.00084	0.00068
		zero and	0.00050			7	0.99906	0	0.98567	0.00086	0.00043
		are not	0.00000			8	0.99950	0	0.99157	0.00057	0.00039
		given.				9	0.99997	0	0.99922	0.00007	0.00004
		The prob-				10	0.99993	0	0.99847	0.00008	0.00007
		lem is				12	0.99995	0	0.99862	0.00006	0.00004
		scaled so				14	0.99999	0	0.99958	0.00001	0.00001
		the vari-				16	0.99999	0	0.99970	0.00001	0.00001
		ance is 1.				20	1.00000	0	0.99992	0.00000	0.00000
						24	1.00000	0	0.99997	0	0
						28	1.00000	0	0.99999	0	0
						32	1.00000	0	1.00000	0	0
						36	1.00000	0	1.00000	0	0

Table 2 (contd)

Name of distribution	Equation	Remarks	First 10 Fourier coefficients	Scale factor	Degree of polynomial for reference	N or number of interpolation points	Area of unnormalized distribution	Mean	Variance	$L_2(-\infty, \infty)$ norm of distribution	$L_\infty(-\infty, \infty)$ norm of distribution
Extreme-	$\frac{1}{\beta}\exp(-y-e^{-y})$		0.54675	1	44	72	1[a]	0.45005[a]	1[a]	0.56625	0.47182
value	$y = x/\beta$		−0.01154								
(Fisher–	$\beta = 0.77969680\ldots$		0.11041								
Tippett			−0.00680			2	0.98605	0.18216	0.96682	0.10732	0.09696
Type I or			−0.06823			3	0.86619	0.41599	0.09617	0.13543	0.12287
doubly			0.01860			4	0.95852	0.44955	0.46446	0.06852	0.06896
exponen-			0.03681			5	1.00335	0.37539	1.05668	0.03311	0.03226
tial)			0.00485			6	0.99298	0.33656	0.96130	0.03167	0.02641
			0.05202			7	0.98698	0.39996	0.80530	0.02521	0.02094
			−0.00680			8	0.98915	0.47332	0.78691	0.01708	0.01180
						9	0.99462	0.43703	0.88470	0.01311	0.01405
						10	0.99696	0.40499	0.96062	0.01083	0.00781
						12	0.99776	0.45280	0.92991	0.00587	0.00549
						14	0.99842	0.43815	0.95782	0.00439	0.00406
						16	0.99938	0.44531	0.97969	0.00274	0.00285
						20	0.99978	0.44498	0.99354	0.00135	0.00121
						24	0.99988	0.44690	0.99594	0.00068	0.00039
						28	0.99992	0.44863	0.99621	0.00036	0.00029
						32	0.99995	0.44962	0.99671	0.00020	0.00019
						36	0.99996	0.45003	0.99755	0.00011	0.00011
						40	0.99998	0.45012	0.99842	0.00006	0.00005

[a]Derived from theoretical calculations.

VI. Checks of Convergence on Standard Probability Density Functions

We now use our method to calculate approximations to some standard probability density functions whose properties (moments, etc.) are known. We chose various one-dimensional distributions to see how fast the method converges. Since different uses may be made of the probability distribution, we will report on a number of measures of how the approximation fits the distribution, including the error in the first two moments, the L_2 $(-\infty, \infty)$ norm, and the uniform norm $[L_\infty(-\infty, \infty)]$. It is hoped that it will be possible to obtain analytic results which will characterize the convergence of this method for different classes of functions.

In describing the rate of convergence of the method, it is important to establish whether the different errors converge to zero as N, the number of interpolation points, goes to infinity. Thus in our numerical study it is desirable to use as high a value for N as possible. But here a difficulty with our method of computation was encountered in that machine overflow occurred for values of N as low as 26. This was because the nature of growth of the value of high-degree polynomials, etc., and the limited range allowed for the magnitude of number on the IBM 7094 (10^{38} in double precision). The problem was reprogrammed using various artifices to keep the numbers within bounds. Now overflow has not occurred for N less than 42.

We list the results for various distributions in Table 2. In the first row we give the name of the distribution, its equation when it is simple, remarks, and its first ten Fourier coefficients. The scale factor refers to the fact that the unit used for the interpolation points may not correspond to the unit for the independent variables of the distribution. We found that generally it is best to scale the problem so that one unit of the scale for the interpolation points corresponds to one standard deviation of the distribution. The scale factor s is defined as follows:

$$P(x) = s\, p(ys)$$

where $P(x)$ is the distribution in terms of the scale for the interpolation points, and $p(z)$ is the usual definition of the distribution.

In order to calculate the L_2 norm of the error of the approximation we needed a reference approximation for the distribution in which the Fourier coefficients are accurately known. A compromise solution was found in which a fixed number of interpolation points (72) was used, and the Fourier coefficients were calculated until machine overflow resulted. Thus in the first example the first 60 coefficients are known. The rest were arbitrarily set equal to zero. It can be shown that, for those examples where the coefficients converge reasonably rapidly, this results in a good approximation in $L_2(-\infty, \infty)$. The L_2 norm of the error of the approximations is obtained from

$$\left(\sum_{i=0}^{71}(a_i - \bar{a}_i)^2\right)^{1/2}$$

where $\bar{a}_i$ is the coefficient from the reference calculation. The moments were obtained by calculating the moments of the approximating function, which was normalized to have unit area.

This study was limited to an examination of differentiable density functions and included others besides those listed here. The origin was chosen to be the mode of the density (all our distributions were unimodal) since it was thought that in applications a first approximation would be available from a maximum likelihood estimate which would be close to the mode of the distribution (we do not assume that the maximum likelihood estimate is calculated correctly). Our findings indicate that for these distributions all of the error norms that we considered converge to zero as N approaches infinity. In addition, it appears that the choice of a scale factor equal to the standard deviation is about as good as can be made.

For smooth functions it was felt that there was no point in using anything but approximations which interpolate the unknown functions. In examples where truncation of the series was tried and weighed mean-square discrete approximation was obtained, the results were always worse than if the interpolated approximation was used (both approximations using the same number of function evaluations).

VII. The Numerical Evaluation of the Coefficients of the Gram–Charlier Expansion

We next show how to evaluate integrals such as (6b) by an approximate numerical method. A Monte Carlo method has already been given (Ref. 7). Any integral of the form

$$\int_a^b y(x)\, w(x)\, dx \tag{40}$$

where $w(x)$ is a weight function can be evaluated by means of a Gaussian quadrature. Here $[a, b]$ may be a finite or infinite interval of the real axis. A weight function $w(x)$ must be such that the moments

$$\int_a^b x^n\, w(x)\, dx = \mu_n \qquad (n = 0, 1, 2, \cdots)$$

exist and are finite for each n, and $w(x) \geq 0$ on $[a, b]$. For any weight function one can construct a corresponding sequence of orthogonal polynomials. Then one has the theorem due to Gauss and Jacobi, which is proved in most tests on numerical analysis:

Theorem. Let $w(x)$ be a weight function for the interval $[a, b]$. There exist real numbers $x_1, x_2, \cdots, x_N, W_1, W_2, \cdots, W_N$ having the properties

1) $a < x_1 < x_2 < \cdots < x_N < b$

2) $W_\nu > 0 \; (\nu = 1, 2, \cdots, N)$

3) The formula $\displaystyle\int_a^b y(x)\, w(x)\, dx = \sum_{\nu=1}^{N} W_\nu\, y(x_\nu)$

is true for every polynomial $y(x)$ of degree $\leq 2N-1$.

Incidentally, the x_ν are the N zeroes of the Nth-degree orthogonal polynomial determined by the weight function $\phi_N(x)$, and the W_ν can also be obtained by various methods from the associated polynomials. These quantities are tabulated for many of the classical polynomials.

Unfortunately, the Gaussian quadrature formulas are not given for $He_n(x)$, but for $H_n(x)$. By a suitable transformation we can make use of the published formulas. Let $\pi_n(\xi)$ be a polynomial of degree n or less.

$$\int_{-\infty}^{\infty} \pi_n(\xi)\, e^{-\xi^2}\, d\xi = \sum_{i=1}^{m} W_i^m\, \pi_n(\xi_i^m) \tag{41}$$

where $2m-1 \leq n$, and the W_i^m and ξ_i^m are tabulated. Now, let $\xi = x/(2)^{1/2}$. The (41) becomes equal to

$$\frac{1}{(2)^{1/2}} \int_{-\infty}^{\infty} \pi_n\left[\frac{x}{(2)^{1/2}}\right] e^{-x^2/2}\, dx$$

Define a new nth-degree polynomial $\phi_n(x) = \pi_n(x/(2)^{1/2})$. Then,

$$\int_{-\infty}^{\infty} \phi_n(x)\, e^{-x^2/2}\, dx = (2)^{1/2} \sum_{i=1}^{m} W_i^m \,\phi(2^{1/2}\xi) \tag{42}$$

Finally, we give a formal derivation of the result. We shall use

$$a_m = \frac{1}{m!} \int_{-\infty}^{\infty} p(x)\, He_m(x)\, dx$$

$$= \lim_{N\to\infty} \frac{1}{m!} \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{1/2}}\, e^{-x^2/2}\, \pi_N(x)\, He_m(x)\, dx$$

where $\pi_N(x)$ is an Nth-degree polynomial in x which is part of the approximation of the integrand.

$$a_m = \lim_{M\to\infty} \lim_{N\to\infty} \frac{1}{m!} \frac{2^{1/2}}{(2\pi)^{1/2}} (2)^{1/2} \sum_{i=1}^{M} W_i^M \pi_N\,[(2^{1/2})\, \xi_i^N]\, He_m\,[(2^{1/2})\, \xi_i^M]$$

$$= \lim_{M\to\infty} \frac{1}{m!} (2)^{1/2} \sum_{i=1}^{M} W_i^M\, p[(2)^{1/2}\, \xi_i^M] \exp\,[(\xi_i^M)^2]\, He_m\,[(2)^{1/2}\, \xi_i^M] \tag{43}$$

This result can be justified in the same manner as for the Hermite series.

For the n-dimensional case, the products $\exp(-x_1^2/2)$ $He_{j_1}(x_1) \exp(-x_2^2/2)\, He_{j_2}(x_2), \cdots$ are used as before. The method of calculating the corresponding coefficients is the same as for the Hermite case.

VIII. An Optimal Property of the Gram–Charlier Expansion

Theorem. Assume that $p(x)$ is given exactly as a combination $\sum_{i=0}^{n} a_i \phi_i(x)$, or alternatively as a combination $e^{-x^2/2}\left[\sum_{i=0}^{n} b_i\, He_i(x)\right]$. Then the approximation obtained by using n interpolation points based on the Gram–Charlier expansion (43) is exact as far as the area and moments up to the $(n-1)$th order.

Proof. Since the method based on the Gram–Charlier expansion is an interpolation using the zeroes of $He_n(x)$, the result will be exact as far as the first n coefficients go, $(b_0, b_1, \cdots, b_{n-1})$. These determine the first $n-1$ moments of the distribution, and the first $n-1$ moments of the distribution (and the area if the distribution is not normalized) determine these coefficients.

IX. Checks on Convergence on Various Probability Density Functions for the Gram–Charlier Approximation

In Table 3 we give some results which correspond to the cases that have previously been done by the Hermite series. It can be seen that the mean and variance converge more rapidly, but that mean-square convergence or uniform convergence need not occur even for smooth functions for which all the moments exist. For instance, for the extreme value distribution using 40 interpolation points, the area and mean are correct to five decimal places for the Gram–Charlier, but off in the fifth and fourth place, respectively, for the Hermite series. However, the L_2 norm and L_∞ norm of the error of the Gram-Charlier approximation diverge in all except one case.

X. Scaling for the *k*-Dimensional Probability Density Function

It is easily verified that if the distribution is Gaussian and is scaled so that the mean is zero and the variance covariance matrix is the unit matrix, then all the coefficients for the Gram–Charlier expansion or the Hermite series expansion are zero except for $a_{00}, \cdots$ or $b_{00}, \cdots$, which equal 1. This suggests scaling the variables so that in the new variables the mean is zero and the variance

Table 3. Convergence of Gram–Charlier approximation

Name of distribution	Scale factor	Number of interpolation points	Area	Mean	Variance	L_2 norm of error	L_∞ norm of error
Unknown phase angle	1	72	0.89464	0.03834	1.00607	0.00000	0.00000
		2	0.89371	0.01270	not defined	0.00637	0.00436
		3	0.89420	0.03808	1.00046	0.00926	0.00724
		4	0.89464	0.03818	1.00405	0.00073	0.00052
		5	0.89464	0.03830	1.00600	0.00026	0.00020
		6	0.89464	0.03834	1.00604	0.00056	0.00049
		7	0.89464	0.03834	1.00606	0.00012	0.00010
		8	0.89464	0.03834	1.00607	0.00003	0.00003
		9	0.89464	0.03834	1.00607	0.00005	0.00005
		10	0.89464	0.03834	1.00607	0.00002	0.00002
		12	0.89464	0.03834	1.00607	0.00001	0.00001
		14	0.89464	0.03834	1.00607	0.00000	0.00000
Cauchy distribution	1	2	0.65774	–	–	0.15407	0.09913
		3	0.82991	–	–	0.14425	0.09250
		4	0.78861	–	–	0.13492	0.09066
		5	0.85464	–	–	0.08738	0.05555
		6	0.84093	–	–	0.12698	0.11324
		7	0.87276	–	–	0.10223	0.06921
		8	0.86864	–	–	0.31294	0.23133
		9	0.88631	–	–	0.12124	0.09026
		10	0.88590	–	–	0.86881	0.76146
		12	0.89780	–	–	3.06090	2.59333
		14	0.90661	–	–	11.80041	10.21137
		16	0.91347	–	–	49.40664	42.85682
		20	0.92359	–	–	0.102×10^{1}	0.895×10^{3}
		40	0.94733	–	–	0.204×10^{11}	0.183×10^{11}
		48	0.95215	–	–	0.250×10^{14}	0.22×10^{14}
Normalized student t distribution, $\nu = 3$	$(3)^{1/2}$	2	0.65774	0	not defined	0.20799	0.23768
		3	1.21284	0	0.36854	0.12301	0.11172
		4	0.84759	0	0.86082	0.18303	0.20991
		5	1.09702	0	0.55811	0.08198	0.06813
		6	0.92305	0	0.82206	0.11766	0.14354
		7	1.04953	0	0.66315	0.06892	0.05695
		8	0.95777	0	0.81389	0.12469	0.14404
		9	1.02707	0	0.72590	0.04485	0.03975
		10	0.97536	0	0.81655	0.03654	0.04593
		12	0.98491	0	0.82311	0.22953	0.22499
		14	0.99038	0	0.83084	0.50863	0.41463
		16	0.99365	0	0.83861	2.02287	1.77571
		48	0.99979	0	0.90469	0.294×10^{12}	0.263×10^{12}
Normalized student t distribution, $\nu = 20$	1.0540	2	0.97285	0	not defined	0.01803	0.01636
		3	1.00220	0	0.92257	0.00974	0.00657
		4	0.99810	0	0.99545	0.01714	0.01545
		5	1.00014	0	0.99057	0.00117	0.00091
		6	0.99979	0	0.99850	0.00123	0.00135
		7	1.00000	0	0.99838	0.00126	0.00090
		8	0.99997	0	0.99955	0.00352	0.00317
		9	1.00000	0	0.99963	0.00015	0.00011
		10	0.99999	0	0.99985	0.00130	0.00104
		12	1.00000	0	0.99995	0.00244	0.00213
		14	1.00000	0	0.99998	0.00285	0.00245
		20	1.00000	0	1.00000	0.01690	0.01478
		36	1.00000	0	1.00000	0.597×10^{2}	0.533×10^{2}
		56	1.00000	0	1.00000	0.533×10^{8}	0.479×10^{8}

Table 3 (contd)

Name of distribution	Scale factor	Number of interpolation points	Area	Mean	Variance	L_2 norm of error	L_∞ norm of error
Extreme value	1	2	0.81657	0.36418	not defined	0.10223	0.09394
		3	1.02431	0.39141	0.53758	0.08334	0.07552
		4	0.98532	0.35252	0.95907	0.10672	0.09395
		5	0.98253	0.47214	0.77955	0.04019	0.03835
		6	1.01088	0.39468	0.96166	0.04312	0.04719
		7	0.98518	0.46418	0.91741	0.07504	0.07114
		8	1.00778	0.42780	0.95980	0.05766	0.05532
		9	0.99343	0.45136	0.97471	0.02371	0.02370
		10	1.00239	0.44489	0.96709	0.06557	0.06054
		12	0.99969	0.45080	0.97845	0.13009	0.10453
		20	0.99991	0.44975	0.99928	6.16576	5.34279
		32	0.99999	0.45007	0.99980	0.922×10^4	0.818×10^4
		40	1.00000	0.45005	0.99998	0.212×10^7	0.190×10^7

covariance matrix is approximately the unit matrix. To illustrate, assume that x is scaled so that the mean is already approximately zero and the variance covariance matrix of x, $E(x\,x^T) \cong \Lambda$ is a positive definite matrix. Let L be a lower triangular matrix such that

$$\Lambda = L\,L^T \tag{44}$$

where L can be determined by the Choleski or square root method. Let X be the transformed variable such that

$$x = LX \tag{45}$$

Then

$$E(x\,x^T) = E(LXX^TL^T) = \Lambda$$

and

$$E(X\,X^T) = L^{-1}\,\Lambda\,(L^T)^{-1} = I \tag{46}$$

Thus the new variables have the required property. The interpolation points are given in terms of the X and transformed by (45) into the old variables.

References

1. Aoki, M., *Optimization of Stochastic Systems*. Academic Press, New York, 1967.

2. Cramér, H., *Mathematical Methods of Statistics*. Princeton Univ. Press, Princeton, N.J., 1951.

3. Kuznetzov, P. I., Stratonovich, R. L., and Tikhonov, V. I., "Quasimoment Functions in the Theory of Random Process," *Dokl. Akad. Nauk SSSR*, Vol. 94, p. 615, 1954. Translation available in *Theory of Prob. and Its Appl.*, Vol. 5, p. 80, 1960.

4. Schwartz, S. C., "Estimation of Probability Density by an Orthogonal Series," *Annals of Mathematical Statistics*, Vol. 38, pp. 1261–1265, 1967.

References (contd)

5. Sansone, G., *Orthogonal Functions.* Interscience Publishers, Inc., New York, 1959.

6. Stroud, A. H., and Secrest, D., *Gaussian Quadrature Formulas.* Prentice–Hall, Inc., N.J., 1966.

7. McShee, R. B., and Walford, R. B., "A Monte Carlo Approach to the Evaluation of Conditional Parameter Estimates for Nonlinear Dynamic Systems," *IEEE Trans. Auto. Control,* Vol. AC-13, p. 29–37, Feb. 1968.